AD________________

Award Number: DAMD17-99-1-9323

TITLE: Antigene Strategy in Breast Cancer Therapy: Rationales for Direct Targeting of erbB2/Her2 DNA with Polyamides

PRINCIPAL INVESTIGATOR: Juan Fernandez Recio, Ph.D.
Vsevolod Katritch, Ph.D.

CONTRACTING ORGANIZATION: The Scripps Research Institute
La Jolla, California 92037

REPORT DATE: September 2002

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 2002	Annual Summary (1 Sep 99 -31 Aug 02)

4. TITLE AND SUBTITLE
Antigene Strategy in Breast Cancer Therapy: Rationales for Direct Targeting of erbB2/Her2 DNA with Polyamides

5. FUNDING NUMBERS
DAMD17-99-1-9323

6. AUTHOR(S):
Juan Fernandez Recio, Ph.D.
Vsevolod Katritch, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

The Scripps Research Institute
La Jolla, California 92037

E-MAIL:
jfrecio@scripps.edu, abagyan@scripps.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
report contains color

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 Words)*

The major goal of this project is to design and computationally evaluate most potent Pyrrole-Imidazole containing polyamide inhibitors of erbB2/Her2 oncogene transcription. We have used an original algorithm to identify the most suitable sequences within erbB2 promoter DNA and then focused our efforts on modeling and design of polyamides with high affinity and specificity to the target these DNA sequences. We have developed a fast and reliable algorithm to build 3-Dimentional molecular models of polyamide-DNA complexes from the corresponding sequences. In our modeling program, PolyGroove, the initial configuration of the complex is generated from standard B-DNA model and the polyamide chain, which is placed in the minor groove according to the specified polyamide-DNA pairing rules. The models are energy optimized with special distance restrains, imposed by the modular nature of polyamide-DNA recognition, and then without any restrains. The algorithm has shown excellent performance in comparative NMR and modeling studies of ten-ring polyamide hairpins, with the control *ab-inito* model closely reproducing all NMR restrains. The PolyGroove program was successfully applied to automatically generate and predict binding energies of polyamide-DNA models with long binding sites (12 and 13 bp)within Erb2/Her2 promoter, using various topologies and a number of new functional groups. Ten most promising candidates for erbB2/Her2 gene-specific inhibition were selected for the further studies.

14. SUBJECT TERMS	15. NUMBER OF PAGES
breast cancer, antigene, polyamides	26
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

Table of Contents

Introduction

Polyamides have been shown to inhibit binding of transcription factors to specific DNA sequences, and thus can be considered as candidate therapeutic agents to regulate gene expression. Pyrrole-Imidazole (Py-Im) containing polyamide molecules can be designed to recognize dsDNA minor groove with high affinity and sequence specificity, comparable to affinity and specificity of gene transcription factors [1-5]. In addition to Pyrrole and Immidazole aromatic rings and their modifications, polyamide chains may contain other "residues" that improve polyamide-DNA specificity [6], interfere with binding of transcription factors[7,8], or enhance cell and nuclear membrane permeability of the polyamide candidate drugs[9]. These common polyamide building blocks, pairing riles and possible topologies are described in **Figures 1-3** and **Table I.**

In this project we design highly specific polyamides to target the erbB2/Her2 promoter region, thus disrupting formation of the transcription complex and inhibiting expression of this important oncogene. The first generation of anti-erbB2 polyamide inhibitors[10], binding DNA sequences in TATAA box proximity, have been shown to effectively inhibit expression erbB2 gene in cell-free expression systems. However, the 7-base pair sequence of polyamide-DNA binding site used in this initial studies is too short and repeats itself ~10^6 times in human genome, questioning safety and efficacy of the candidate drug based on these polyamide constructs.

The major goal of our study is to rationally select longer (12-16 bp) dsDNA targets in erbB2 promoter to achieve maximum whole genome specificity and to design optimal polyamide binders to these regulatory sites.

Body

Task1: Optimization of target sequences in gene Her2/erbB-2 promoter.

The sequence of the erbB2 gene promoter contains well-characterized TATAA and CCAAT boxes, repetitive GGA motif and putative SP1 binding sequences in the region upstream to the major transcription start site, see **Figure 4**. Despite TATA presence, multiple transcription start sites have been found, the major ones being 21 and 70 bp down from the TATA box. It was shown that the 500bp region upstream of the major starting site is sufficient for both basal and inducible transcription activity, the most proximal 125bp DNA stretch being responsible for about 30-fold overexpression in most cancer cell lines [11].

<u>a. List all short (8-16 bp) sequences, flanking TATA , CAAT and GC boxes in Her2 promoter.</u>

We performed a comprehensive database analysis, based on the specialized MatInspector tool [12,13], to find putative regulatory elements in the 500 bp promoter of erbB2. **Table II** lists the results of this search for the most important 150 bp proximal region. Most sites, found and characterized previously, were identified in the search (these entries are emphasized both in **Table II** and **Figure 4**). For example, the ETS response element next to the TATAA box[10,11], as well as AP-2 binding site[14], CCAAT box, were identified.

Based on the analysis presented in **Table II** we selected 6 short 16 bp sequences, flanking

transcription factor binding sites, see **Figure 4**. Note that four of these sequences overlap with more that one major activation site, which makes them the most interesting targets for antigene therapy.

b. Rate specificity of the listed sequences in the human genome.

Published human genome sequence gives us an opportunity to predict the specificity of a polyamide binder on a whole genome level. We have designed a specialized program to perform exhaustive BLAST-based searches in the human genome draft to assign sequence specificity of a particular binding pattern. We performed both searches for exact sequence matches, as well as a simple sequence profile search with low penalty for A-T substitution. The latter approach was devised to take into account full degeneracy of Py-Py recognition of A-T pair and partial degeneracy of Pyrrole-Hydrohypyrrole (Py-Hp) recognition of A-T. Using this program, we assigned the specificity to all possible 11, 12, 13 and 14 bp fragments within preselected target sequences. **Figure 5** demonstrates an example result of our analysis in the case of 13 bp fragments.

c. Rate conservation of listed sequences using several versions of the promoter.

Conservation of the target sequence is critical for development of effective antigene inhibitors. Analysis of the 6 available versions of the erbB2 promoter sequences from different sources have demonstrated good sequence conservation in the chosen proximate region from -150 to 0, while more deletions-insertions are possible in the farther upstream sequence. In the proximate region, the erbB2 promoter sequence can contain gaps in positions -135 and –122 and an A->T mutation in position –69, corresponding sites are shown in **Figure 5** in red.

d. Sort the list of target sequences.

We sorted all the short fragments within highlighted sites based on the sequence specificity score, length and overlap with core activation sites. This analysis has produced several nontrivial insights. First, we found that the whole region around the TATAA box, which is very important for regulation of gene activity, has very poor specificity in the human genome[10]. In addition, sequence **6** is very AT rich, which further lowers its polyamide specificity score. On the other hand, sequences **1**, **2**, and **4** contain 13 bp fragments with almost unique whole-genome specificity, and each of them overlap with more than one activation site.

As a result of the above sequence analysis of erbB2 promoter performed in the Task 1, the sequences in **Table II** have been chosen as optimal targets for polyamide design, see **Table III**. The most promising target is the DNA sequence **4**, which overlaps with 2 important regulatory sites of erbB2/Her2 promoter, is almost unique in the genome, does not have documented variations in the sequence, and also have a low AT content, benefitial for polyamide recognition specificity.

Task 2: Overall design and evaluation of complimentary polyamides.

a. For each target sequence generate a set of polyamide molecules using DNA-polyamide recognition code and a choice of additional blocks.

Using a set of polyamide elements and polyamide-DNA pairing rules [15,16], summarized in

Table 1, we have devised an algorithm to build all matching polyamide sequences for each target dsDNA site. The algorithm starts by building a "perfect match" sequence that contains Py, Im and Hp rings only and performs all possible substitutions and connections to allow various types of topology suggested in the proposal. Additional empirical rules are also applied to eliminate unfeasible designs, e.g. only 2 to 4 successive rings are allowed, β-alanines are isolated, only 4 γ-links are allowed, and so on. With these restrictions applied, the program automatically generates as many as ~30-50 different polyamide sequences for each 13 bp DNA sequence or ~20-30 polyamides for 12 bp DNA. We performed this procedure with the best 50 DNA targets from our target list and stored the resulting 1285 "sequences" of polyamide-DNA complex in a specialized database.

b. Check feasibility of chemical synthesis for designed compounds.

Polyamide chains, containing various combinations of Imidazole (Im), Pyrrole (Py), Hydroxypyrrole (Hp) rings, β-alanin, γ-linkers, and many other building blocks can be produced by Boc solid phase chemistry using standard protocols, described in works from Peter Dervan's laboratory[8,17-21]. Recently, Fmoc solid phase chemistry have been also introduced for a machine-assisted synthesis of Im-Py polyamides[22], as well as oxime resin chemistry, which allows extension of the polyamide C-terminal tails repertoire[23]. In our design we utilize a standard set of residues and overall topologies, with proven chemical feasibility. While some designs here may be preferred over others, currently no theoretical limitations have been found on chemical feasibility of polyamides in our database.

c. Make preliminary estimations for affinity and specificity of each compound.

The central part of our project is 3D modeling of the resulting DNA-polyamide complexes and evaluation of their relative affinity. Our original algorithm uses the fact that polyamide complexes with DNA are very modular in structure. This allows us to build initial conformations of new complexes, based on known X-ray geometries of previously characterized complexes[24-26]. The program tethers DNA and ligand residues to the respective residues in the X-ray structure. These initial conformations are subsequently optimized by restrained energy minimization, where energy terms include bonded, van der Waals, electrostatic and hydrogen bonding terms. The application of geometry restraints enforces DNA-DNA base-pairing and DNA-polyamide pairing rules in the initial stage of the optimization, forcing the model to follow the "canonical" pattern of polyamide-DNA recognition . In the final stage, the restraints are removed and free global energy minimization is applied. The deviation between restrained and free energy minimized models is usually within all-atom RMSD < 1.5 Å for "match" polyamide-DNA complexes, which suggest high quality of the modeling. Single polyamide mismatches increase this RMSD to ~2-3 Å, thus reflecting big deviations of the fully energy-optimized model from the "canonical" recognition pattern.

The polyamide-DNA binding energy of the models was estimated in terms of van der Waals, hydrogen bonding, electrostatic and solvation contributions. Comparison with more than 50 published measurements for short polyamide hairpins estimates the accuracy of relative binding energy predictions at about 1.7 kcal/mol. This polyamide-DNA modeling algorithm was presented at the Program in Mathematics and Molecular Biology meeting.

Task 3: Detailed modeling and selection of candidate structures

a. Test and adjust the ICM global minimization procedure with published polyamide-DNA complexes.

The polyamide modeling algorithm was further upgraded to accommodate new variants of polyamide topology and improve affinity estimations by using a more accurate molecular force field. We have also adjusted the procedure for automated 3-D modeling of polyamide–DNA complexes to making conformational and binding energy predictions more robust for longer complexes with new design elements.

The first improvement deals with the choice of starting configurations of the complex and polyamide placement. The new algorithm uses standard B-DNA as initial conformation, and places the polyamide chain into the DNA minor groove according to the specified polyamide-DNA pairing rules. Only then the special distance constraints, provided by the available polyamide-DNA X-ray structures are employed in the energy optimization of the complex. These modifications help to avoid strong deviations from B-DNA structure in the initial steps of the procedure and provide much faster and more reliable convergence for energy minimizations.

The other improvement takes advantage of the new internal coordinate force field (ICFF) developed in the lab[27]. The ICFF is automatically generated from a "source" Cartesian force field (such as MMFF94s or Amber) with an algorithm that "projects" Cartesian parameters into the torsion coordinate space. Implicit flexibility, naturally incorporated into the torsion energy parameters, is critical to the accuracy of the internal coordinates model with fixed covalent geometry. Essential also is the ability of ICFF method to generate fixed covalent geometries for new chemical structures, using Cartesian geometry minimization with the source force field. This feature facilitates inclusion of new elements into our custom polyamide residue library, producing fixed residue geometries compatible with the new torsion force field. Direct modifications (i.e. aromatic ring to β-alanine replacement) in polyamide chain sequence are now allowed through fast local geometry optimization in Cartesian coordinates, followed by internal coordinate global optimization.

Prediction accuracy of the new algorithm with ICFF geometries and energy functions substantially improved compared to the previous version with ECEPP torsion potential, reducing geometry RMSD from ~1.2 Å to just ~0.9 Å in our standard comparison test with of available PDB entries (365d and 334d). Binding free energy estimations with the new algorithm also improved from 1.7 kcal to 1.3 kcal RMSD.

Prediction power of our polyamide-DNA modeling algorithm was also evaluated in NMR structural study, performed in collaboration with Dr. Wemmer group[28]. A conformational model of 10-ring hairpin-DNA complex, derived by our algorithm *ab-initio* was found to be in excellent agreement with the corresponding NMR model, built with NOESY distance constraints, RMSD < 1 Å (see the poster presentation attached).

b. Build all-atom models for DNA complexes with newly designed polyamides.

The automated procedure for polyamide design was programmed with ICM molecular modeling package, which takes DNA sequences and coded polyamide sequences as input, and produces energy optimized complexes in the output. An example of the program input and output are shown in **Figure 6**.

The program reads the input sequence where each DNA and polyamide "residue" is represented with one letter or symbol. Double stranded DNA is built in a standard energy optimized B-form by an original ICM script. A polyamide chain of specific sequence (or two

chains in case of overlapping hairpin topology) is built from the library of residues. The pairing between polyamide residues and DNA residues is assigned according to the input. One or more X-ray templates are then superimposed onto the DNA structure to cover the polyamide binding site, and the polyamide atoms are "tethered" to the corresponding polyamide atoms in the templates.

Tight binding of polyamides in the DNA minor groove and the modular nature of the pairing between the molecules suggest special approach to energy minimization of the complex. We apply so-called ICM "regularization" procedure to minimize both length of the "tethers" and the conformational energy of the object. Regularization procedure goes through several iteration steps, using different weight ratio for conformational energy and "tether tension" energy at each minimization step. The weight of the tethers in the energy function gradually decreases throughout the regularization procedure, making the final solution virtually independent on the tethers. Minimizations, performed in torsion coordinates, not only guarantee fast convergence of this procedure, but also prevent severe deformations in covalent geometry due to the tether tension in the initial steps of the procedure. Spatial positions of the templates are readjusted in the course of the regularization procedure to allow large-scale movement of DNA backbone. This annealing-like algorithm is designed to generate low-energy structures with high local similarity to the templates.

For each of the three selected 16-bp DNA targets, we generated more than 100 polyamide "perfect match" complexes with 12-bp DNA recognition sites, which differ in positions of 5-member rings in the sequence or in overall topology. We use several criteria to check the quality of the models built. First, we check the length of hydrogen bond contacts between polyamide and DNA residues, which are expected by the pairing rules. For the best models we found up to 93% of the of the 34 hydrogen bonds within 2.5 Å lengths (measured as hydrogen to heavy atom distance), while on the average about 89% of the H-bonds satisfy this criteria for the "perfect match" models. Second, we check the tethers between the model and the template, and found that the average length of the tethers is about 0.5 Å and usually do not exceed 1.5 Å. Finally, we performed 10 independent runs with single mismatches in the polyamide sequences and found the consistent increase in the complex conformational energy compared to the perfect match case.

A new important polyamide residue, *N*-diaminoalkylpyrrole, have been added recently to the polyamide design repertoire [8]. Polyamides with diaminoalkyl "positive patch" not only allow reliable inhibition of transcription factors with exclusive major groove binding, e.g. bZIP proteins, but also improve affinity and specificity of DNA recognition. Thus, using alkylpyrrole positive patch in combination with C-terminal N-methylamide as a "tail" we might be able to improve polyamide gene inhibitors in many cases (**Figure 7**). We designed and optimized geometry of new *N*-diaminoalkylpyrrole, *N*-diaminoalkylimidazol and N-methylamide residues, and incorporated them into the library of polyamide elements.

c. <u>Calculate global minimum conformations for each complex and evaluate polyamide-DNA binding energy.</u>

The annealing procedure, employed in the global energy optimization of the complex is described above. We performed a separate study with three polyamide-DNA complexes to assess global convergence of energy optimizations in our special case. For each model we used 20 independent runs of the procedure with different annealing schedules. In all the three cases we found slight variability in the results of different runs, with the average conformational energy RMSD ~0.7 kcal and geometry RMSD~0.9 Å. Such conformational variability is expected in the polyamide-DNA complexes, and has to be taken into account by

averaging results over several independent runs.

Much more flexible aminoalkyl and C-terminal methylamide moieties of polyamides were treated separately with the ICM Monte Carlo global optimization method to allow large-scale changes in their conformations. ICM allows freezing of the variables in the rest of the complex, which makes exhaustive Monte Carlo search in the flexible parts of the molecule possible on a reasonable time scale. We found this Monte-Carlo search critical to avoid local minima trapping of the flexible parts of the polyamide molecule.

Polyamide-DNA binding energy for a given conformation of the complex was predicted as a sum of hydrogen bonding, van der Waals and electrostatic interactions energies between polyamide and DNA, combined with different weights (1., 0.43 and 0.75 respectively). This binding energy formula was previously found to be optimal by calibration with shorter polyamides[29]. For each polyamide-DNA complex, the binding energy was calculated as an average of binding energies of five independently minimized conformations. Binding energy results for the best polyamide binders to the erbB2 promoter sequence **4** are presented in **Table IV.** Note, that affinity of the "tandem hairpin" design in our predictions is consistently better, compared to single-molecule topologies, i.e. soft hairpin and cyclic chains. These results can be explained by somewhat higher conformational flexibility of the tandem hairpin topology, as well as better affinity of newly discovered optimal short tails to the G•C base pair[23] ("-" = $NH(CH_2)_2OH$ tail, "~"= $NH(CH_3)$ tail). Also, our results confirm that the novel positively charged diaminoalkyl extensions tend to improve overall DNA binding affinity of polyamides in addition to their role in enhancing interference with the gene transcription[8].

To represent diversity of the polyamide topology, five best "tandem haipins", three "soft haipins" and two "cyclic polyamides" in **Table IV** have been selected for as lead erbB2 inhibitors for future investigations. Structure of the best tandem hairpin complex is presented in **Figure 8.**

Task 4. *In vitro* and *in vivo* testing

a. Test designed polyamide compounds *in vitro* for their DNA sequence specificity and ability to block transcription factors binding to erbB2/Her2 promoter.

b. Test these compounds for their efficacy in human breast cancer cell cultures.

The experimental testing is not budgeted in the current grant and is expected to be performed through an academic collaboration. Recently published data indicate that with the exception of certain T-cell lines, polyamide-dye conjugates tend to localize mainly in the cytoplasm, but not in the nucleus of live cells [9,30]. Specifically, the study from Peter Dervan's group arrived to the conclusion that previously designed 8-ring polyamides[10], though very strong erbB2 inhibitors in cell-free expression systems, may be not effective against breast cancer cell lines due to their inability to access nuclear DNA[9]. These new circumstances make our potential collaborators to postpone synthesis and testing of novel anti-erbB2 polyamides until the problem of cell nucleus delivery of polyamides is solved.

Several groups are currently working on possibility to design new generation of polyamide-like molecules with improved nuclear localization[9,31] and we plan to provide our expertise in computer-assisted polyamide design to these groups to facilitate development of polyamide conjugates with nuclear localization, without sacrificing their DNA binding affinity and specificity.

Key Research Accomplishments

- found the most important candidate targets for antigene therapy within the proximal erbB2 promoter
- estimated the whole-genome specificity of all possible short fragments within this promoter region
- designed an automatic algorithm to list all possible polyamide topologies matching a given DNA sequence
- written a program, generating initial 3D models of a polyamide-DNA complex from its "sequence", based on the known pattern of polyamide-DNA recognition and global energy optimization in torsion coordinates
- employed a novel accurate force field (ICFF) in the modeling algorithm, making feasible reliable calculations for longer polyamide-DNA complexes and facilitating new design topologies
- benchmarked and optimized our predictions of polyamide-DNA binding affinity, using available experimental data
- tested the quality of our 3D models in a joint modeling-NMR study of 10 ring polyamide hairpins, complexed with DNA
- included new aminoalkyl-modified residues in the polyamide residue library, improving both affinity and inhibitory effect of the designed polyamides
- generated all-atom models for more than 300 polyamides complexed with DNA targets in erbB2 gene promoter
- predicted binging energy of these polyamides and selected most potent polyamide designs for further experimental studies

Reportable outcomes

- Programs and algorithms:
  - PolyVar program to generate possible polyamide sequences for a given DNA recognition site.
  - PolyGroove© program for fast 3D modeling of polyamide-DNA complexes from the corresponding residue sequences and subsequent binding affinity predictions (requires ICM-pro package).
  -
- Meeting Presentation and Abstracts:
  - Katitch, V., Abagyan, R.A. and Olson, W.K. (1999). Structural Modeling of Polyamide-DNA Recognition. *Mathematics and Molecular Biology VI*, Santa Fe, NM
  - M. Totrov, V. Katritch, D. Pilch,* W.K. Olson,* J. Fernandez-Recio, R. Abagyan, Flexible Docking (2000). *The Scripps Research Institute Scientific report*, La Jolla, CA.
  - Bernhard H. Geierstanger, Colin J. Loweth, Vsevold Katritch, Ruben Abagyan, Peter G. Schultz & David E. Wemmer (2001). NOE distance constraints and structural modeling of a ten-ring hairpin complex with DNA. *Frontiers of NMR and Molecular Biology Meeting*, Keystone, CO.
  - Vsevold Katritch, Juan Fernandez Recio and Ruben Abagyan (2002) Targeting of erbB2/Her2 DNA with polyamides. *Era of Hope Department Of Defense (DOD)Breast Cancer Research Program (BCRP) meeting*, Sept 24-28, Orlando, FL.

- Articles:
  - Vsevolod Katritch, Maxim Totrov and Ruben Abagyan (2002). ICFF: A new method to incorporate implicit flexibility into an internal coordinate force field. *J. of Comp. Chem. in press.*
  - The modularity of DNA recognition by polyamide molecules persists for a ten-ring hairpin in complex with an eight base pair binding site. Bernhard H. Geierstanger, Colin J. Loweth, Vsevold Katritch, Ruben Abagyan, Peter G. Schultz & David E. Wemmer. (2002) Submitted to *J. of Am. Chem. Soc.*

Conclusions

In this project, we have identified the best candidate dsDNA targets for polyamide binding within the most important proximal region of the erbB2 promoter sequence and sorted them according to their whole-genome specificity and overlap with transcription activation sites. Using an extended set of binding blocks, choice of topology variants and an original automated procedure, we have listed chemically feasible polyamides matching the target dsDNA sequences, according to the polyamide-DNA pairing rules. We have developed a fast and reliable algorithm to build 3D models of these polyamides-DNA complexes, based on the known modular structure of the complexes and all-atom conformational energy minimization. The accuracy of our structural modeling were confirmed by experimental NOESY distance constraints, and binding energy predictions were extensively benchmarked with available data on short polyamide hairpin-DNA affinity.

Using these algorithms, we have build more than 300 polyamide-DNA models targeting 12 and 13 base pair recognition sites within the three selected erbB2 promoter targets. Analysis of polyamides DNA hydrogen bonding pattern and energy strain in the complex suggests that even for such extended complexes all specific polyamide-DNA contacts can be conformationally afforded, if we use optimal polyamide chain topologies, with no more than 4 aromatic rings in a row. Also our modeling suggests that diaminoalkyl group conjugated to an aromatic residue not only extend the molecule into DNA major groove but also can substantially improve polyamide-DNA binding affinity. Binding energy evaluations allowed selection of the best candidates for each of the 3 best topologies, including tandem hairpins, soft hairpins and cyclic chains.

The 10 chosen polyamide structures are expected to have high binding affinity and whole genome specificity to the erbB2 promoter DNA and can be considered as highly specific erbB2 inhibitors with potential anti-cancer activity. Further development of these lead candidates for breast cancer drug requires optimization of nuclear membrane permeability of polyamide-like molecules and further study of pharmacokinetic features of polyamides.

	G•C	C•G	T•A	A•T
I/P, I/β	+	-	-	-
P/I, β/I	-	+	-	-
H / P	-	-	+	-
P / H	-	-	-	+
P/P, β/P, P /β	-	-	+	+
γ-linker (R)^{H2N} γ-linker	-	-	+	+
β, β/β	-	-	+	+

Table I. Polyamide-DNA pairing rules. Along with Pyrrole (P), Imidazole (I) and Hydrohypyrrole (H) rings, other elements include β-alanine, which can stack with any ring or with itself to provide some flexibility, as well as two types of γ-links, used as flexible "connectors" linking opposite polyamide strands.

Name of family/matrix	Further Information	Position	Strand	Core sim.	Matrix sim.	Sequence
V$SP1F/GC_01	GC box elements	-148:-135	(+)	0.876	0.790	gctgGGAGttgccg
V$LYMF/TH1E47_01	Thing1/E47 heterodimer	-134:-119	(-)	1.000	0.910	aacgaagtCTGGgagt
V$CMYB/CMYB_01	c-Myb	-120:-103	(+)	1.000	0.949	ttggaatgcaGTTGgagg
V$VMYB/VMYB_02	v-Myb	-113:-105	(-)	0.819	0.899	tccAACTgc
V$COMP/COMP1_01	COMP1	-89:-66	(-)	1.000	0.781	tcctgtgATTGggagcaagcgcgc
V$PCAT/CAAT_01	cellular and viral CCAAT box	-82:-71	(+)	1.000	0.890	tgctcCCAAtca
V$ECAT/NFY_01	nuclear factor Y (Y-box binding factor)	-82:-67	(+)	1.000	0.920	tgctcCCAAtcacagg
V$VDRF/VDR_RXR_B	VDR/RXR heterodimer site	-69:-55	(+)	1.000	0.906	aggagaagGAGGagg
V$VDRF/VDR_RXR_B	VDR/RXR heterodimer site	-57:-43	(+)	1.000	0.892	aggtggagGAGGagg
V$AP2F/AP2_Q6	activator protein 2	-51:-40	(-)	0.857	0.772	agCCCTcctcct
V$ETSF/ETS1_B	c-Ets-1 binding site	-36:-22	(+)	1.000	0.910	tgaGGAAgtataaga
V$TBPF/TATA_C	Retroviral TATA box	-30:-21	(+)	0.843	0.779	agTATAAGAa
V$NFKB/NFKB_Q6	NF-kappaB	-8:-5	(-)	1.000	0.830	agGGGAatctcagc
V$NOLF/OLF1_01	olfactory neuron-specific factor	-1:-20	(-)	1.000	0.822	ctccggTCCCaatggaggggaa

Table II. Results of MatInspector analysis for 600 bp promoter fragment containing the major transcriptional start site (position 0), CCAAT and TATAA boxes, ETS response element and other potential targets for antigene therapy.

No	DNA Sequence	Regulatory elements
1	[illegible]GTTGCCGACTCCCAG	GC box element and Thing1/E47 heterodimer
2	CTTCGTTGGAATGCAG	c-Myb
4	GAGCGCGCTTGCTCCC	COMP1 and CCAAT box

Table III. erbB2 promoter sites selected for polyamide targeting. Regulatory elements, possibly involved in erbB2 activation are highlighted. Documented single nucleotide polymorphism (SNP) sites are shown in red.

<table>
<tr><th>Input sequence</th><th>Topology type</th><th>Predicted binding energy, kcal</th></tr>
<tr><td>GAGCGCGCTTGCTCCC
IPIbIP-hIK
+ +
PIp~PIbPPI
CTCGCGCGAACGAGCC</td><td>Tandem soft hairpins,
with g-NH_3^+ linkers,
with diaminoalkyl group</td><td>23.2±0.9</td></tr>
<tr><td>GAGCGCGCTTGCTCCC
IPIbIP-pIK
+ +
PIp~PIbPPI
CTCGCGCGAACGAGCC</td><td>Tandem soft hairpins,
with g-NH_3^+ linkers,
with diaminoalkyl group</td><td>21.5±1.4</td></tr>
<tr><td>GAGCGCGCTTGCTCCC
IPIbIP-pIK
+ +
PIp-PIbPPI
CTCGCGCGAACGAGCC</td><td>Tandem soft hairpins,
with g-NH_3^+ linkers,
with diaminoalkyl group</td><td>-21..3±1.6</td></tr>
<tr><td>GAGCGCGCTTGCTCCC
IPI-iPbPIK
+ +
PIPbPi-PPI
CTCGCGCGAACGAGCC</td><td>Tandem soft hairpins,
with g-NH_3^+ linkers,
with diaminoalkyl group</td><td>-121.0 ±1.2</td></tr>
<tr><td>GAGCGCGCTTGCTCCC
IPI-iPbHIP
+ +
PIPbPi-PPI
CTCGCGCGAACGAGCC</td><td>Tandem soft hairpins,
with g-NH_3^+ linkers,
no diaminoalkyl group</td><td>-20.1±1.6</td></tr>
<tr><td>GAGCGCGCTTGCTCCC
iPIbIPbPIK
|
~PIPbPIbPPI
CTCGCGCGAACGAGCC</td><td>Soft hairpin,
with $NH(CH_3)$ tail and
with diaminoalkyl group</td><td>-19.4 ± 2.5</td></tr>
<tr><td>GAGCGCGCTTGCTCCC
IPIPbPPbIP~
|
PIPIbIPbPi
CTCGCGCGAACGAGCC</td><td>Soft hairpin (reverse strand),
with $NH(CH_3)$ tail and
no diaminoalkyl group</td><td>-18.5 ± 1.7</td></tr>
<tr><td>GAGCGCGCTTGCTCCC
iPIPIPbPIK
+
-IPPIPIbPPI
CTCGCGCGAACGAGCC</td><td>Soft hairpin,
with $(CH_2)_2OH$ tail and
with diaminoalkyl group</td><td>-18.2 ± 1.7</td></tr>
<tr><td>GAGCGCGCTTGCTCCC
iPIbIPbPIK
+ +
IPbIPIbPPI
CTCGCGCGAACGAGCC</td><td>Soft cyclic,
with g- and g-NH_3^+ linkers,
with diaminoalkyl group</td><td>-17.3 ± 1.4</td></tr>
<tr><td>GAGCGCGCTTGCTCCC
IPIbIPbHIK
+ |
PIPbPIbPPI
CTCGCGCGAACGAGCC</td><td>Soft cyclic,
with g- and g-NH_3^+ linkers,
with diaminoalkyl group</td><td>-16.9 ± 1.5</td></tr>
</table>

Table IV. Top ten suggested polyamide binders to the erbB2 promoter target sequence **4**. Accuracy of the energy predictions was assessed by five independent annealing minimizations. One-letter codes for polyamide residues are: "P"- pyrrole, "I"- Imidazole, "H"- hydroxypyrrole, K- diaminoalkylpyrrole, R- diaminoalkylimidazole, "b"- β-alanine, "|"- γ-linker, "+" - γ-NH_3^+ linker, "_" β-DP tail, "-"-$NH(CH_2)_2OH$ tail, "~"-$NH(CH_3)$ tail[23]. The second polyamide molecule is colored red.

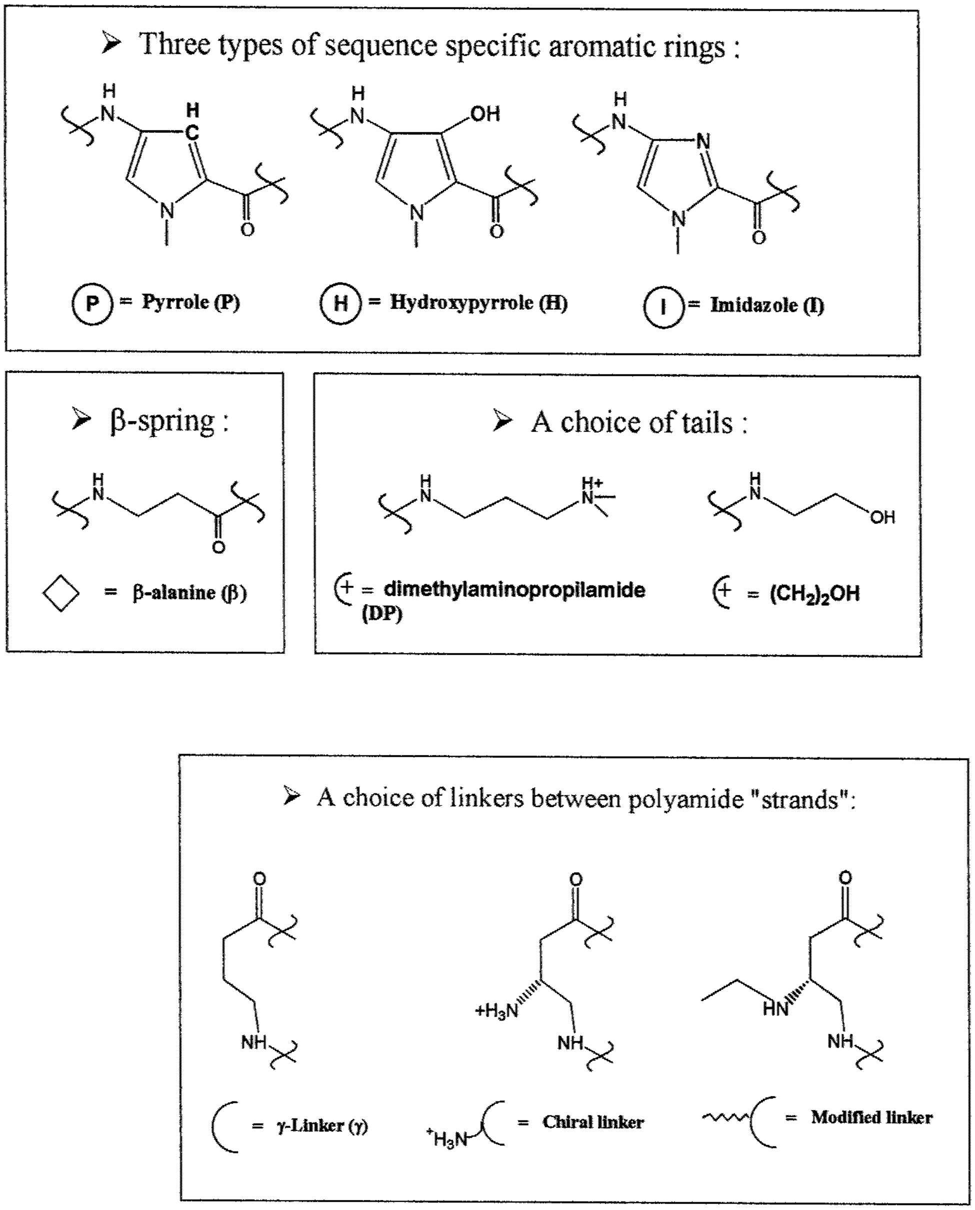


Figure 1. Aromatic and aliphatic residues, employed in design of highly specific DNA ligands[16].

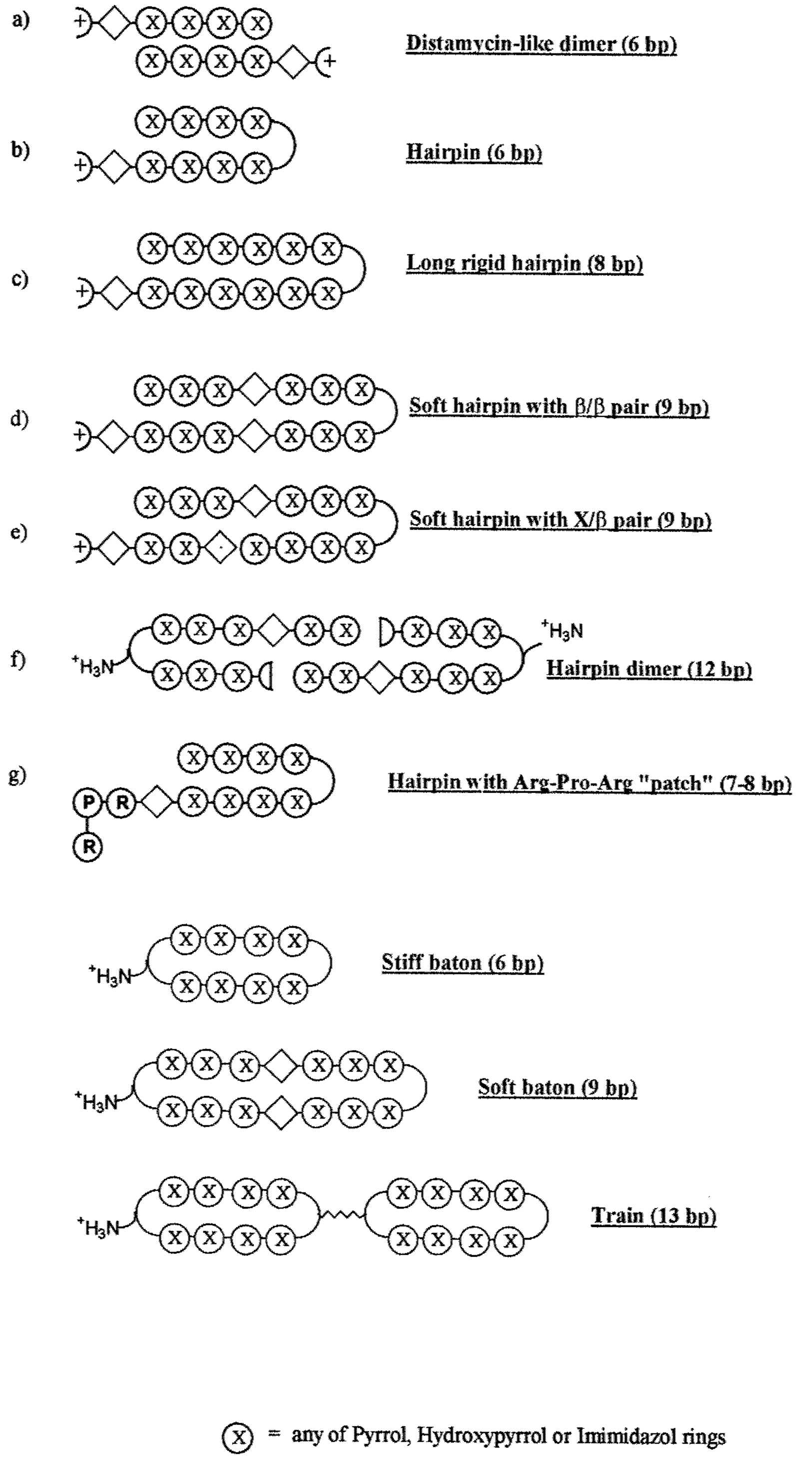


Figure 2. Various topologies used in polyamide design[16].

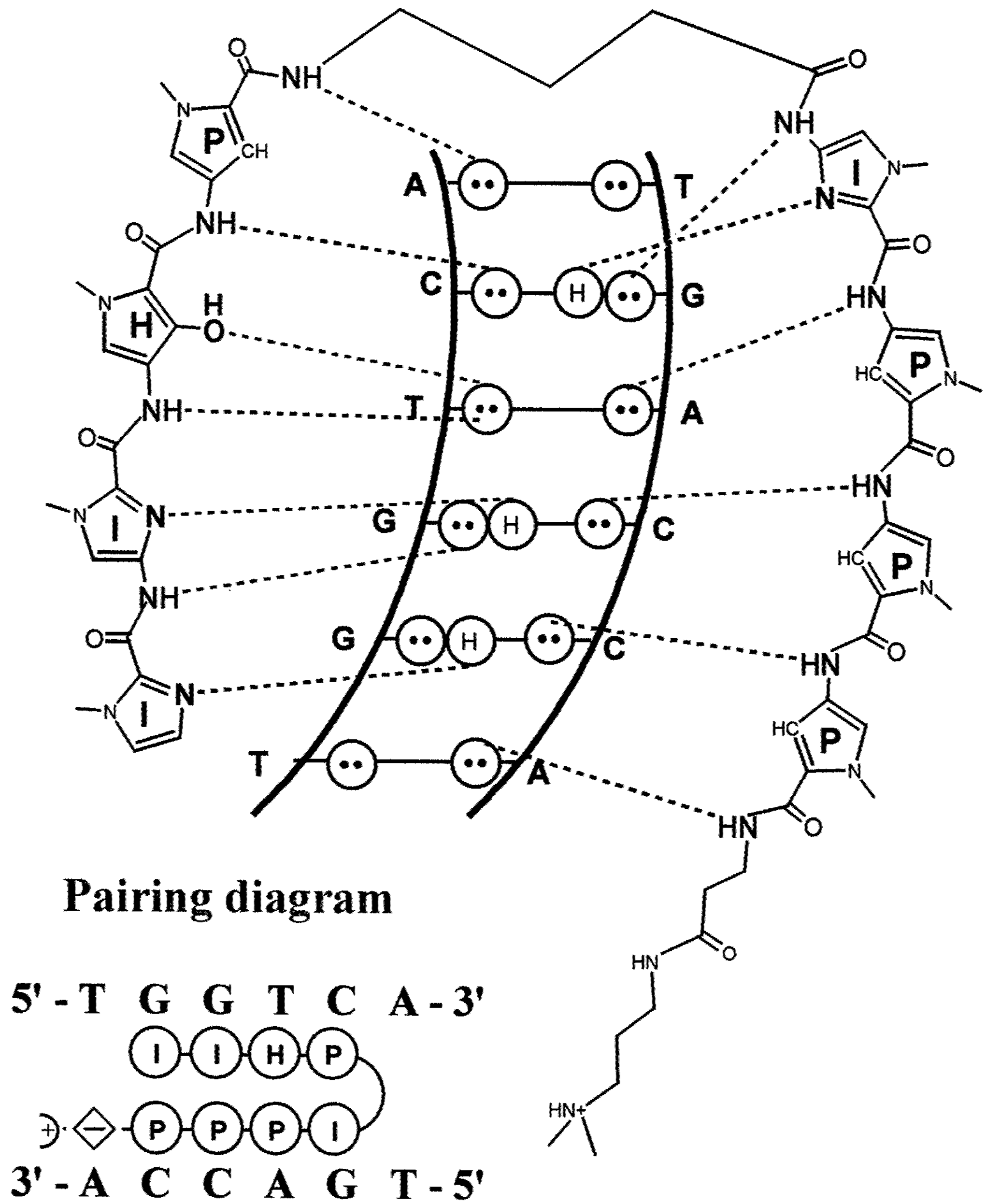


Figure 3. Structural basis of polyamide-DNA recognition. Hydrogen bonds, required for binding specificity of Pyrrole (P), Imidazole (I) and Hydroxypyrrole (H)are shown as dashed lines. Also shown standard diagram presentation of the complex.

```
                1                       2                 3
-150> AGCTGGGAGTTGCCGACTCCCAGACTTCGTTGGAATGCAGTTGGAGGGGG

           ------>  4   <------
-100> TGAGCTGGGAGCGGCGCTTGCTCCCAATCACAGGAGAAGGAGGAGGTGGAG

              5                   6
-50 > GAGGAGGGCTGCTTGAGGAAGTATAAAGAATGAAGTTGTGAAGCTGAGATT<0
```

Figure 4. Sequence of the proximal region of erbB2 promoter. Predicted core activation sites are underscored and experimentally confirmed sites are shown in **bold**. Also the arrows show two palindromic sequences[10] involved in transcription activation. We have highlighted and numbered 16-bp sequences, chosen as putative targets for further analysis.

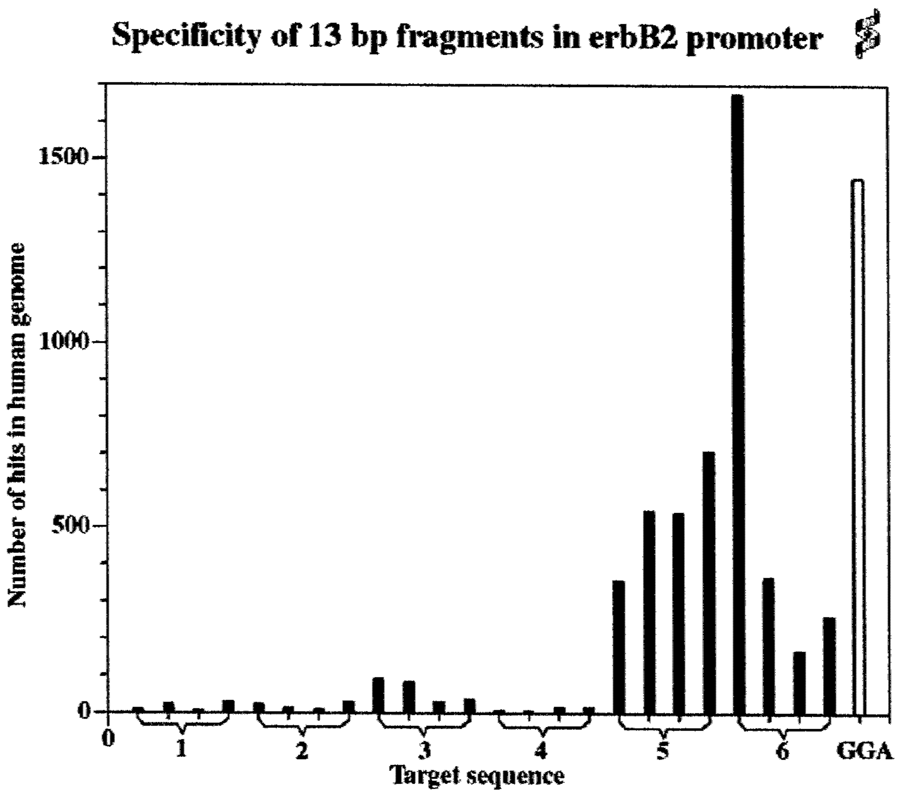


Figure 5. Whole-genome specificity analysis for 13 bp fragments of the proximal erbB2 promoter sequence. Note that the most rare fragments correspond to sequences **1**, **2** and **4** respectively (see Figure 1), while fragments **5** and **6** flanking TATA box have very poor whole-genome specificity, comparable to the specificity of the control fragment with a GGA repeat (the white bar).

INPUT SCRIPT:

```
#!/home/sevak/icm2/icmL
call _startup
call _PolyGroove  ## Polyamide modeling tools in ICM scripting language

#aregul  (template_obj)(DNA_seq)(Polyamide_seq)(i_start)(n_steps)(l_display)(l_freeMin)

aregul "hp1_template.ob" "GGGAGCGCGCTTGCTCCCA" "IPI-iPbPIP+IPP-iPbPIP+" 5 100 no no

quit
```

OUTPUT FILE:

```
GGGAGCGCGCTTGCTCCCA+IPI-iPbPIP+IPP-iPbPIP+.ob
#_summary   : icmName        GGGAGCGCGCTTGCTCCCA+IPI-iPbPIP+IPP-iPbPIP+
#_summary   : objCode        hp1_template.ob
#_summary   : nChains        4
#_summary   : chainList      watson crick a b
#_summary   : nResidues      60
#_summary   : nFreeVar       322
#_summary   : vwCutoff       7.5
#_summary   : hbCutoff       3.0
#_summary   : electroMethod  distance dependent
#_summary   : dielConst      4.0
#_summary   : surfaceMethod  atomic solvation
#_summary   : eTotal         -1208.74
#_summary   : grad           290.42
#_summary   : eVacuum        -917.15
#_summary   : eNonEl         -695.44
#_summary   : e_vw           -751.66
#_summary   : e_hb           -79.73
#_summary   : e_to           135.95
#_summary   : e_el           -221.71
#_summary   : eSolvat        -291.59
#_summary   : eEntropy       0.00
#_summary   : tzWeight       0.24
#_summary   : rmsd           1.00
#_summary   : rmsdBackbone   1.04
#_summary   : nTz            320
#_summary   : resNotTz       18
```

Figure 6. PolyGroove input and output files for one of the DNA-polyamide sequences. "connectors" linking opposite polyamide strands.

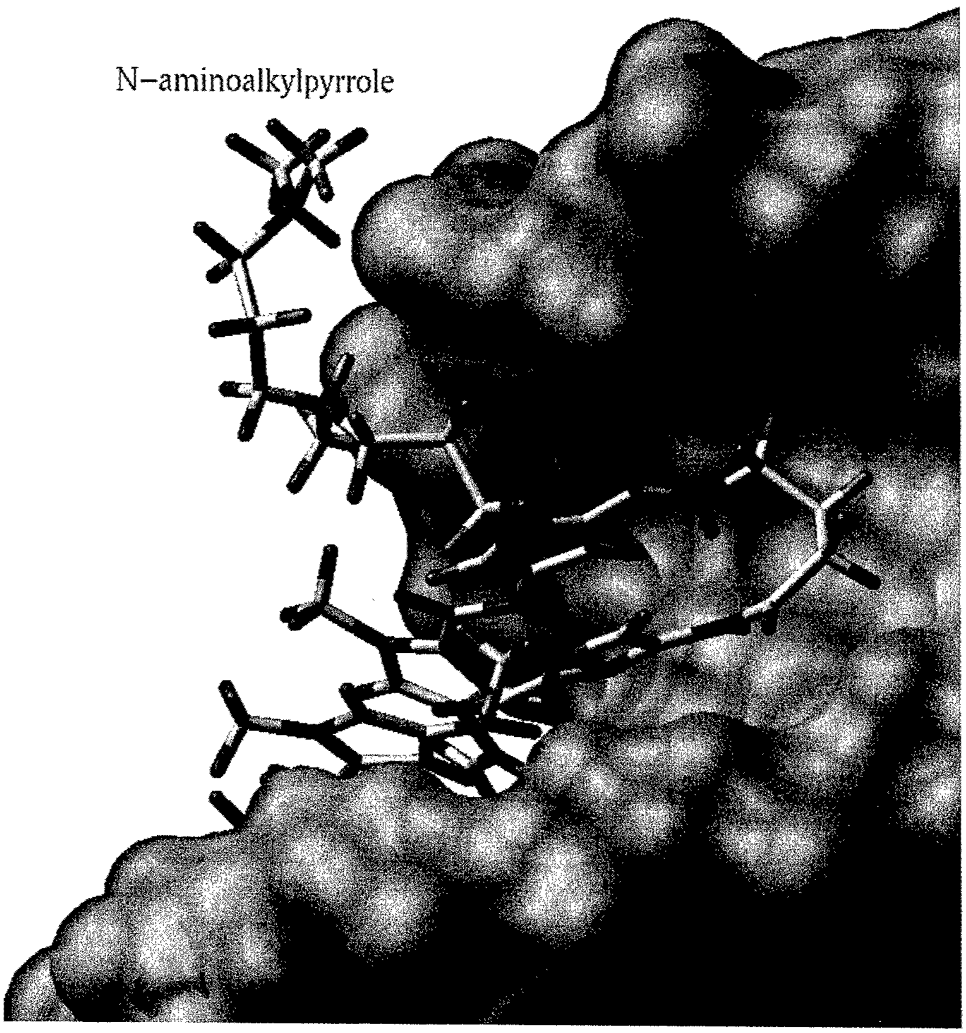


Figure 7. N-diaminoalkylpyrrole containing polyamide in the DNA minor groove. This globally optimized conformation shows interaction of the diaminoalkyl tail with the DNA phosphates, which ensures inhibition of major-groove binding transcription factors by polyamides of this type.

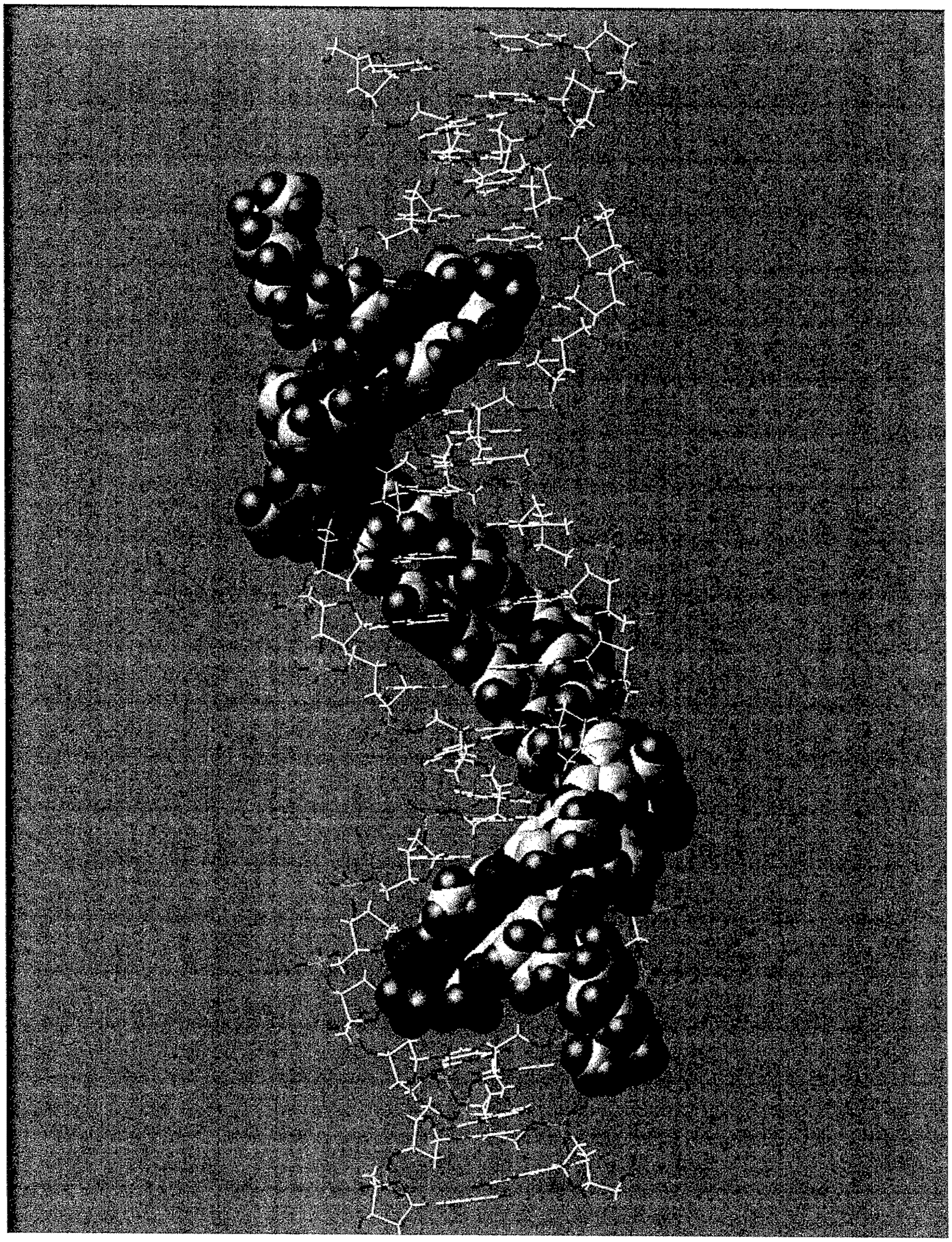


Figure 8. Recognition of the target erbB2/Her2 DNA sequence **4** by the 8-ring tandem hairpin polyamide, predicted to have the best binding energy among ~300 polyamide designs tested. Pairing diagram is shown below:

```
GAGCGCGCTTGCTCCC
   IPIbIP-hIK
  +          +
   PIp~PIbPPI
CTCGCGCGAACGAGCC
```

References:

1. Lown, J.W. DNA recognition by lexitropsins, minor groove binding agents. *J Mol Recognit* **7**, 79-88. (1994).
2. White, S., Baird, E.E. & Dervan, P.B. On the pairing rules for recognition in the minor groove of DNA by pyrrole-imidazole polyamides. *J Mol Biol* **274**, 439-45. (1997).
3. Kielkopf, C.L., Baird, E.E., Dervan, P.B. & Rees, D.C. Structural basis for G.C recognition in the DNA minor groove. *Nature* **391**, 468-71. (1998).
4. White, S., Szewczyk, J.W., Turner, J.M., Baird, E.E. & Dervan, P.B. Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Chem Biol* **5**, 119-33. (1998).
5. Wemmer, D.E. & Dervan, P.B. Targeting the minor groove of DNA. *Chem Biol* **4**, 569-78. (1997).
6. Wang, C.C., Ellervik, U. & Dervan, P.B. Expanding the recognition of the minor groove of DNA by incorporation of beta-alanine in hairpin polyamides. *Bioorg Med Chem* **9**, 653-7. (2001).
7. Bremer, R.E., Baird, E.E. & Dervan, P.B. Inhibition of major-groove-binding proteins by pyrrole-imidazole polyamides with an Arg-Pro-Arg positive patch. *Science* **282**, 111-5. (1998).
8. Bremer, R.E., Wurtz, N.R., Szewczyk, J.W. & Dervan, P.B. Inhibition of major groove DNA binding bZIP proteins by positive patch polyamides. *Bioorg Med Chem* **9**, 2093-103. (2001).
9. Belitsky, J.M., Leslie, S.J., Arora, P.S., Beerman, T.A. & Dervan, P.B. Cellular uptake of N-methylpyrrole/N-methylimidazole polyamide-dye conjugates. *Bioorg Med Chem* **10**, 3313-8. (2002).
10. Chiang, S.Y. *et al.* Targeting the ets binding site of the HER2/neu promoter with pyrrole- imidazole polyamides. *J Biol Chem* **275**, 24246-54. (2000).
11. Scott, G.K. *et al.* Binding of an ETS-related protein within the DNase I hypersensitive site of the HER2/neu promoter in human breast cancer cells. *J Biol Chem* **269**, 19848-58. (1994).
12. Quandt, K., Frech, K., Karas, H., Wingender, E. & Werner, T. MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res* **23**, 4878-84. (1995).
13. Quandt, K., Grote, K. & Werner, T. GenomeInspector: a new approach to detect correlation patterns of elements on genomic sequences. *Comput Appl Biosci* **12**, 405-13. (1996).
14. Bosher, J.M., Williams, T. & Hurst, H.C. The developmentally regulated transcription factor AP-2 is involved in c-erbB-2 overexpression in human mammary carcinoma. *Proc Natl Acad Sci U S A* **92**, 744-7. (1995).
15. Wemmer, D.E. & Dervan, P.B. Targeting the minor groove of DNA. *Curr Opin Struct Biol* **7**, 355-61. (1997).
16. Dervan, P.B. Molecular recognition of DNA by small molecules. *Bioorg Med Chem* **9**, 2215-35. (2001).
17. Baird, E.E. & Dervan, P.B. Solid Phase Synthesis of Polyamides Containing Imidazole and Pyrrole Amino Acids. *J. Am. Chem. Soc.* **118**, 6141-6146 (1996).

18. Parks, M.E., Baird, E.E. & Dervan, P.B. Optimization of the Hairpin Polyamide Design for Recognition of the Minor Groove of DNA. *J. Am. Chem. Soc.* **118**, 6147-6152 (1996).
19. Turner, J.M., Swalley, S.E., Baird, E.E. & Dervan, P.B. Aliphatic/Aromatic Amino Acid Pairings for Polyamide Recognition in the Minor Groove of DNA. *J. Am. Chem. Soc.* **120**, 6219-6226 (1998).
20. Trauger, J.W., Baird, E.E. & Dervan, P.B. Recognition of 16 Base Pairs in the Minor Groove of DNA by a Pyrrole-Imidazole Polyamide Dimer. *J. Am. Chem. Soc.* **120**, 3534-3535 (1998).
21. Herman, D.M., Turner, J.M., Baird, E.E. & Dervan, P.B. Cycle Polyamide Motif for Recognition of the Minor Groove of DNA. *J. Am. Chem. Soc* **121**, 1121-1129 (1999).
22. Wurtz, N.R., Turner, J.M., Baird, E.E. & Dervan, P.B. Fmoc solid phase synthesis of polyamides containing pyrrole and imidazole amino acids. *Org Lett* **3**, 1201-3. (2001).
23. Belitsky, J.M., Nguyen, D.H., Wurtz, N.R. & Dervan, P.B. Solid-phase synthesis of DNA binding polyamides on oxime resin. *Bioorg Med Chem* **10**, 2767-74. (2002).
24. Kielkopf, C.L. *et al.* A structural basis for recognition of A.T and T.A base pairs in the minor groove of B-DNA. *Science* **282**, 111-5. (1998).
25. Kielkopf, C.L., Baird, E.E., Dervan, P.B. & Rees, D.C. Structural basis for G.C recognition in the DNA minor groove. *Nat Struct Biol* **5**, 104-9. (1998).
26. Kielkopf, C.L. *et al.* Structural effects of DNA sequence on T.A recognition by hydroxypyrrole/pyrrole pairs in the minor groove. *J Mol Biol* **295**, 557-67. (2000).
27. Katritch, V., Totrov, M. & Abagyan, R. ICFF: A new method to incorporate implicit flexibility into an internal coordinate force field. *J. Comp. Chem.* **in press**(2002).
28. Geierstanger, B. *et al.* The modularity of DNA recognition by polyamide molecules persists for a ten-ring hairpin in complex with an eight base pair binding site. *J. of Am. Chem. Soc* **Submitted**(2002).
29. Katrich, V., Abagyan, R.A. & Olson, W.K. Structural modeling of polyamide-DNA recognition. in *Abstracts of PMMB VI, Santa Fe, New Mexico* (, 1999).
30. Sharma, S.K., Morrissey, A.T., Miller, G.G., Gmeiner, W.H. & Lown, J.W. Design, synthesis, and intracellular localization of a fluorescently labeled DNA binding polyamide related to the antibiotic distamycin. *Bioorg Med Chem Lett* **11**, 769-72. (2001).
31. Burli, R. *et al.* DNA binding ligands with excellent antibiotic potency against drug-resistant gram-positive bacteria. *Bioorg Med Chem Lett* **12**, 2591. (2002).

Bibliography:

- Meeting Presentation and Abstracts:
  - Katitch, V., Abagyan, R.A. and Olson, W.K. (1999). Structural Modeling of Polyamide-DNA Recognition. *Mathematics and Molecular Biology VI*, Santa Fe, NM
  - M. Totrov, V. Katritch, D. Pilch,* W.K. Olson,* J. Fernandez-Recio, R. Abagyan, Flexible Docking (2000). *The Scripps Research Institute Scientific report*, La Jolla, CA.
  - Bernhard H. Geierstanger, Colin J. Loweth, Vsevold Katritch, Ruben Abagyan, Peter G. Schultz & David E. Wemmer (2001). NOE distance constraints and structural modeling of a ten-ring hairpin complex with DNA. *Frontiers of NMR and Molecular Biology Meeting*, Keystone, CO.
  - Vsevold Katritch, Juan Fernandez Recio and Ruben Abagyan (2002) Targeting of erbB2/Her2 DNA with polyamides. *Era of Hope Department Of Defense (DOD)Breast Cancer Research Program (BCRP) meeting*, Sept 24-28, Orlando, FL.

- Articles:
  - Vsevolod Katritch, Maxim Totrov and Ruben Abagyan (2002). ICFF: A new method to incorporate implicit flexibility into an internal coordinate force field. *J. of Comp. Chem. in press.*
  - The modularity of DNA recognition by polyamide molecules persists for a ten-ring hairpin in complex with an eight base pair binding site. Bernhard H. Geierstanger, Colin J. Loweth, Vsevold Katritch, Ruben Abagyan, Peter G. Schultz & David E. Wemmer. (2002) *Prepared for submission.*

Personnel list:

1. Vsevold Katritch	PI (09/1999-11/2001)
2. Juan Fernandez Recio	PI (11/2001-08/2002)

The modularity of DNA recognition by polyamide molecules persists for a ten-ring hairpin in complex with an eight base pair binding site

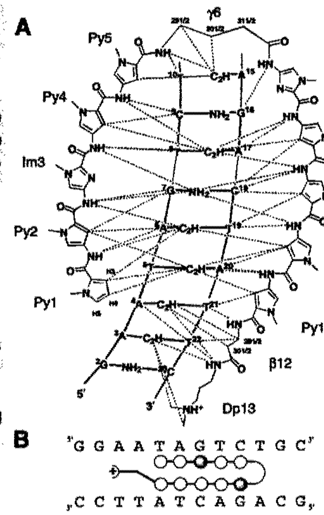
Bernhard H. Geierstanger (1), Vsevold Katritch (2), Colin J. Loweth (2), Ruben Abagyan (1,2), Peter G. Schultz (1,2) & David E. Wemmer (3)

(1) Genomics Institute of the Novartis Research Foundation, 3115 Merryfield Row, San Diego, CA 92121-1125
(2) Department of Chemistry and Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Rd., La Jolla, CA 92037
(3) Department of Chemistry, University of California, Berkeley, CA 94720

Summary

Polyamides containing imidazole (Im), pyrrole (Py) and hydroxypyrrole (Hp) building blocks recognize DNA through specific contacts in the minor groove and can interfere with gene expression (Ref. 1). The most studied polyamide ligands consist of three or four Py/Im residues linked via a hairpin residue to a second set of three or four rings followed by a tail. Here we present the first structural data on the complex of a ten-ring polyamide with a 8 base pair target site, using 2D NOESY data combined with restrained molecular modeling. The high modularity of polyamide-DNA complexes allowed us to develop a computer script for the molecular modeling program ICM (Ref. 2) to quickly generate starting models for NMR refinements from the geometry of polyamide residues in previously studied complexes. This was applied to the ten-ring hairpin ligand Py-Py-Im-Py-Py-γ-Im-Py-Py-Py-Py-β-Dp bound to d(GGAATAGTCTGC)*d(GCAGACTATTCC). NOE-restrained molecular modeling with ICM indicates a complex consistent with the rules discovered previously. Broadening of NMR resonance lines of the first and the tenth ring residue that are stacked on top of each other indicate conformational exchange in this part of the complex. This in turn suggests energetically unfavorable contacts with the DNA as is expected for a ligand of this size.

A



B

5'G G A A T A G T C T G C3'
3'C C T T A T C A G A C G5'

C

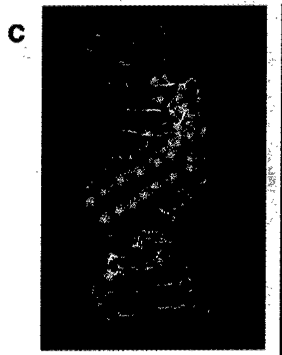
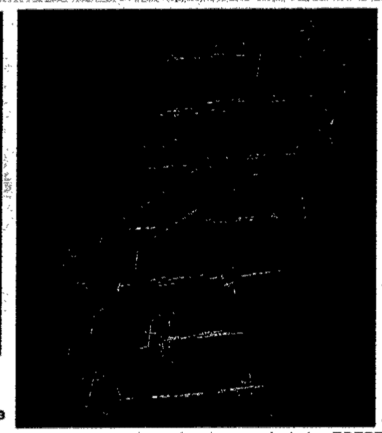
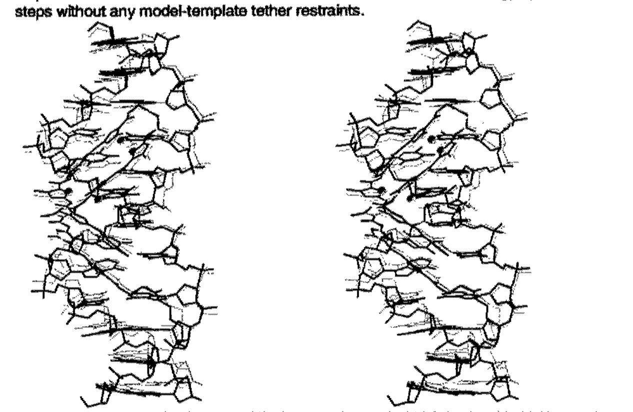


Fig. 1. (A) Structure of the polyamide hairpin Py-Py-Im-Py-Py-γ-Im-Py-Py-Py-Py-β-Dp. NOEs to H1' and adenine H2 protons in the minor groove of d(GGAATAGTCTGC)*d(GCAGACTATTCC). (B) Schematic representation of the ten-ring hairpin complex indicating orientation and residue stacking. Shaded circles represent N-methylimidazole ring residues; open circles for N-methylpyrrole rings. (C) Molecular model with hairpin ligand in green.



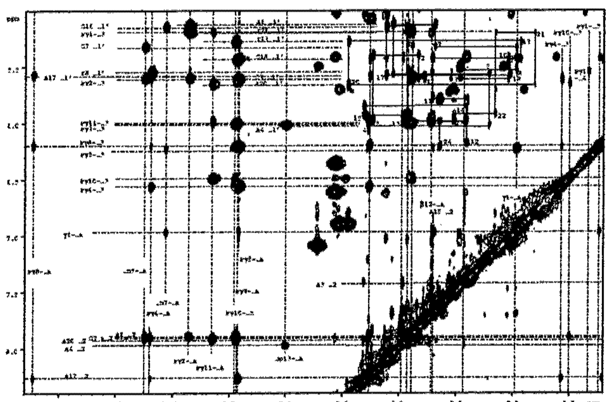
Molecular Modeling with ICM (Molsoft)

Fig. 4. An automatic procedure for the ICM molecular modeling package was developed to quickly build and energy-optimize a molecular model of any polyamide-DNA complex of interest (Ref. 2). Starting conformations for the DNA and for the polyamide ligand are generated from a library of standard geometries derived from published X-ray data of polyamide-DNA (PDB: 365D, 407D and 408D) complexes. The ICM script tethers DNA and ligand residues to the respective residues in a X-ray template (407D) and overlays model and template by minimizing the length of the tethers. This is followed by an energy optimization procedure using internal coordinates with fixed geometry and free torsional angles. The all-atom conformational energy includes ECEPP/3 terms (van der Waals, electrostatic, hydrogen bonding and torsional energy) plus an harmonic term for model-template tether restraints. The polyamide linker and tail residues are optimized using ICM's Monte Carlo global energy optimization procedure. To avoid local energy minima the strength of the model-template tethers is changed every 1000 steps during a total of 20000 minimization steps. A final restraint-free model is obtained after another 10000 energy optimization steps without any model-template tether restraints.



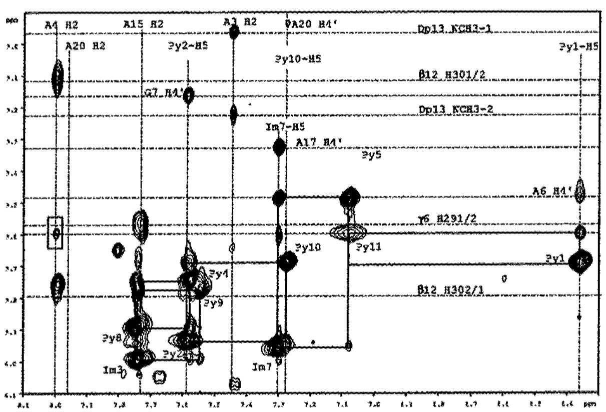
NOE contacts verify binding in minor groove, suggest hydrogen bonds to G amino groups and line broadening of selective ligand resonances indicates conformational exchange of terminal pyrroles

Fig. 2. 2D NOESY (95% H_2O/5% D_2O, 400 MHz, 25 C, τ_{mix} = 200 ms). Sequential aromatic to H1' connectivities for the DNA duplex are shown as solid lines with nucleotide numbers indicating the intraresidue aromatic to H1' cross-peaks. Dashed lines indicate resonance lines of ligand amide and pyrrole protons, and of DNA protons in NOE contact with ligand protons (Fig. 1A for labeling). Chemical shift values and NOE contacts of the G7 and G16 amino groups suggest hydrogen bonds to imidazole Im3 and Im7. Py2, β12 and Dp13 amide as well as Py1 and Py11 proton resonances are broadened by conformational exchange.



Upfield shifted pyrrole H5 resonances and line broadening indicate conformational exchange and unusual stacking of the terminal ligand pyrrole rings Py1 and Py11

Fig. 3. 2D NOESY (in 100% D_2O, 400 MHz, 25 C, τ_{mix} = 200 ms). N-methylpyrrole or imidazole H5 to N-methyl proton connectivities characteristic for the residue stacking arrangement shown in Fig. 1B are drawn as solid squares. Ring residue numbers indicate the intraresidue N-methyl proton to H5 cross-peaks. Dashed lines indicate resonance lines of ligand or DNA protons in NOE contact (Fig 1A for labeling). H5 proton resonances of Py11 and Py1 are broadened by conformational exchange. The unusual chemical shift of Py1-H5 suggests unusual stacking interactions with Py11.



Restraint-free model and NMR-restrained model generated with ICM overlay well, and DNA binding specificity is determined by ligand imidazole hydrogen bonds with guanine amino groups

Fig. 5. Stereo diagram of Py-Py-Im-Py-Py-γ-Im-Py-Py-Py-Py-β-Dp in complex with d(GGAATAGTCTGC)*d(GCAGACTATTCC). Overlayed are the models optimized with (black lines) and without semiquantitative distance restraints derived from NOE data (gray lines) (RMSD for all atoms: 1.0 Å). In the NMR refined model 80 ligand-DNA, 45 intramolecular ligand-ligand restraints from 100 ms NOE data as well as 30 DNA base pairing restraints were used in the ICM energy optimization procedure described in Fig. 4. Hydrogens have been omitted for clarity. Hydrogen bonded imidazole ligand nitrogens and guanine amino nitrogens important for the sequence specificity of the hairpin-DNA complex are shown as gray spheres.

References:

(1) (a) Wemmer, D. E., Annu Rev Biophys Biomol Struct 2000, 29, 439-61. (b) Dervan, P.B.; Burli, R. W., Curr Opin Chem Biol. 1999, 3, 688-93. (c) Gottesfeld, J.M., Turner, J.M., Dervan, P.B., Gene Expr 2000, 9, 77-9.
(2) (a) Abagyan, R.A.; Totrov, M.M.; Kuznetsov, D.A. J. Comp. Chem. 1994, 15, 488-506. (b) Katritch, V.; Abagyan R. A.; Olson, W. K. Abstracts of PMMB VI, Santa Fe, New Mexico 1999.
(c) Nemethy, G.; Gibson, K.D.; Palmer, K.A.; Yoon, C.N.; Paterlini, G.; Zagari, A.; Rumsey, S. & Scheraga, H.A. J. Phys. Chem. 1992, 96, 6472-6484.

THE SCRIPPS RESEARCH INSTITUTE

Targeting erbB2/Her2 DNA with polyamides

Vsevold Katritch (1,2), Juan Fernandez Recio (1) and Ruben Abagyan (1)

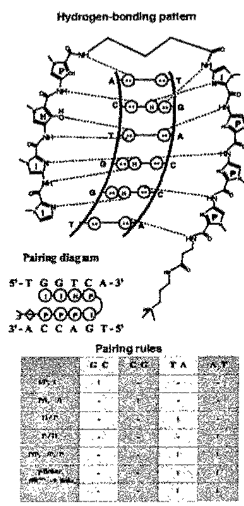
(1) Department of Chemistry and Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Rd., La Jolla, CA 92037

(2) Plexus Vaccine Inc., 11770 Bernardo Plaza Ct #370B, San Diego, CA 92128

Summary

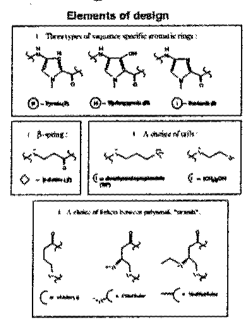
Tyrosine kinase receptor erbB2/HER2/neu oncogene, a key component in the epidermal growth factor (EGF) signaling pathway, is amplified and upregulated in 25-30% of human breast cancers and is associated with poor clinical prognoses. Specific inhibition of the gene on the transcriptional level (antigene strategy) would have a high therapeutic potential. We suggest using a novel class of Pyrrole-Imidazole (Py-Im) containing polyamides to bind specific DNA sequences in the erbB2 promoter region to disrupt transcription. We have applied sequence analysis tools to identify the most promising short targets within erbB2 DNA promoter sequence and performed extensive molecular modeling to design optimal polyamide molecules that bind these dsDNA targets. We found that the region around the TATAA box, previously used as an antigen target (Chiang, S.Y. et al. J Biol Chem 275, 24246-54. (2000)), has very poor specificity in the human genome. On the other hand, we discovered sequences containing 13 bp fragments with almost unique whole-genome specificity, also overlapping with one or more erbB2 activation sites. As a result of our analysis, several short fragments erbB2 promoter sequences have been chosen as optimal targets for polyamide design, including GC and CCAAT boxes, c-Myb and AP2 regulatory sites.

A fast and reliable algorithm have been developed with ICM modeling software to build 3D models of polyamides-DNA interaction, based on the known modular structure of the complexes and all-atom conformational energy minimization with ICFF potential function (Katritch, V., Totrov, M. & Abagyan, R., J. Comp. Chem. (2002)). The affinity of the DNA - polyamide binding can be predicted by this method with an accuracy of ~1.5 Kcal/mol, which significantly narrows the search for the best candidate polyamides for future in vitro and in vivo experiments, the accuracy of such modeling demonstrated in NMR experiments (Gelerstanger, B., Katritch, V., et al. J. of Am. Chem. Soc Submitted(2002)). Using polyamide virtual "residue" library we have generated more than a 100 different polyamide "sequences" for each of the dsDNA binding sites, build 3D models of these molecules and selected a few best designs according to their predicted binding affinity.



Intro: Polyamide-DNA recognition

Elements of design



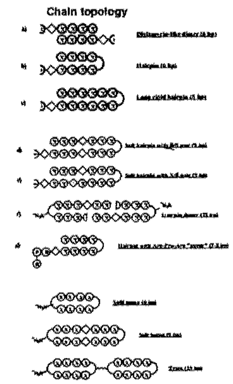
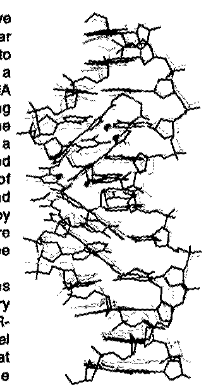


Fig.1. Heteroaromatic polyamide molecules can be designed to recognize dsDNA minor groove with high affinity and sequence specificity, comparable to affinity and specificity of gene transcription factors[1]. In addition to Pyrrole (Py), Immidazole (Im) and Hydroxypyrrole (Hp) rings responsible for DNA specificity, described by a set of "pairing rules", polyamides may contain "tails", "linkers" and other "residues" that allow various topology types of the polyamide chain, directly interfere with binding of transcription factors, or add newfunctional features to the molecules.

Structural Modeling with ICM (Molsoft)

Fig. 2. An automatic PolyGroove procedure for the ICM molecular modeling package was developed to quickly build and energy- optimize a molecular model of any polyamide-DNA complex of interest (Ref. 2). Starting conformations for the DNA and for the polyamide ligand are generated from a library of standard geometries derived from published X-ray data of polyamide-DNA (PDB: 365D, 407D and 408D) complexes. This is followed by by an energy optimization procedure using internal coordinates with free torsional angles.

Comparative NMR-modeling studies confirm high accuracy of geometry predictions, with RMSD between NMR-restrained (black) and *ab-initio* model for a 10 ring hairpin complex less that 1.0 A. Also, for short hairpins the algorithm can predict binding energies with accuracy about 1.5 Kcal [2].



Analysis of the erbB2/Her2 promoter sequence points to target site #4 as overlapping with most important regulatory sites, most specific in human genome and having the highers GC content.

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
V$SP1F/GC_01	GC box elements	-148;-135	(+)	0.876	0.790	gctgGGAGtgcag
V$LYMF/TH1E47_01	Thing1/E47 heterodimer	-134;-119	(-)	1.000	0.910	aacgaagCTGGgagt
V$CMYB/CMYB_01	c-Myb	-120;-103	(+)	1.000	0.949	tggaagcaGTTGgagg
V$VMYB/VMYB_02	v-Myb	-113;-105	(-)	0.819	0.899	tccAACTgc
V$COMP/COMP1_01	COMP1	-89;-68	(-)	1.000	0.781	tccagagATTGgagcagagcgc
V$PCAT/CAAT_01	cellular and viral CCAAT box	-82;-71	(+)	1.000	0.890	tgctcCCAAtca
V$ECAT/NFY_01	nuclear factor Y (Y-box binding factor)	-82;-67	(+)	1.000	0.920	tgctcCCAAtcacaca
V$VDRR/VDR_RXR	VDR/RXR heterodimer site	-69;-55	(+)	1.000	0.906	aggaggagGAGGagg
V$VDRR/VDR_RXR	VDR/RXR heterodimer site	-57;-43	(+)	1.000	0.892	aggaggagGAGGagg
V$AP2F/AP2_04	activator protein 2	-51;-40	(-)	0.857	0.772	agCCCTcctcct
V$ETSF/ETS1_B	c-Ets-1 binding site	-36;-22	(+)	1.000	0.910	tgaGGAAgtataaga
V$TBPF/TATA_C	Retroviral TATA box	-30;-21	(+)	0.843	0.779	agTATAAGAa
V$NFKB/NFKB_Q6	NF-kappaB	-6;-5	(-)	1.000	0.830	agGGGAatctcagc
V$NOLF/OLF1_01	olfactory neuron-specific factor	-1;-20	(-)	1.000	0.822	ctcaggTCCCaaggdeggggaa

Table I. Results of MatInspector analysis for 600 bp promoter fragment containing the major transcr iptionaltar site (position 0). CCAAT and TATAA boxes .ETS response element and other potential targets for antigene therapy

No	DNA Sequence	Regulatory elements
1	[redacted]TTGCCGACTC[redacted]	GC box element and Thing1/E47 heterodimer
2	CTTCGTTGGAATGCA[redacted]	c-Myb
4	[redacted]CGCGCTTGCTC[redacted]	COMP1 and CCAAT box

Table II. erbB2 promoter sites selected for polyamide targeting. Regulatory elements possib ly involved in erbB2 activation are highlighted. Documented single nucleotide polym

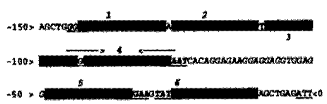


Fig. 3. Sequence of the proximal region of erbB2 promoter. Predicted core activation sites are underscored and experimentally confirmed sites are shown in bold. Also the arrows show two palindromic sequences10 involved in transcription activation. We have highlighted and numbered 16-bp sequences, chosen as putative targets for further analysis

Specificity of 13 bp fragments in erbB2 promoter

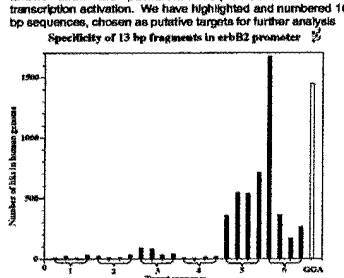


Fig. 4. Whole-genome specificity analysis for 13 bp fragments of the proximal erbB2 promoter sequence. Note that the most rare fragments correspond to sequences 1, 2 and 4 respectively (see Figure 1), while fragments 5 and 6 flanking TATA box have very poor whole-genome specificity, comparable to the specificity of the control fragment with a GGA repeat (the white bar)

Rational design of polyamides, targeting site 4 (GAGCGCGCTTGCTCCC) . Structural modeling, analysis of conformational strain and binding energy evaluation can lead to potent candidate anti-erbB2 ligands.

Input sequence	Topology type	Predicted binding energy, kcal
[illegible]	Tandem soft hair pins, with g-NH3+ linkers, with diaminoalkyl group	-23.1+0.9
[illegible]	Tandem soft hair pins, with g-NH3+ linkers, with diaminoalkyl group	-21.5+1.4
[illegible]	Tandem soft hair pins, with g-NH3+ linkers, with diaminoalkyl group	-21.3+1.6
[illegible]	Tandem soft hair pins, with g-NH3+ linkers, with diaminoalkyl group	-21.0+1.2
[illegible]	Tandem soft hair pins, with g-NH3+ linkers, no diaminoalkyl group	-20.1+1.6
[illegible]	Soft hair pin, with NH(CH3) tail and with diaminoalkyl group	-19.4+2.5
[illegible]	Soft hair pin (reverse strand), with NH(CH3) tail and no diaminoalkyl group	-18.5+1.7
[illegible]	Soft hair pin, with $(CH_2)_2OH$ tail and with diaminoalkyl group	-18.2+1.7
[illegible]	Soft cyclic, with g- and g-NH3+ linkers, with diaminoalkyl group	-17.3+1.4
[illegible]	Soft cyclic, with g- and g-NH3+ linkers, with diaminoalkyl group	-16.9+1.5

Table III. Top ten suggested polyamide binders to the erbB2 promoter target sequence 4. Accuracy of the energy predictions was assessed by five independent annealing minimizations. One-letter codes for polyamide residues are: "P"- pyrrole, "I"- Imidazole, "H"- hydroxypyrrole, K- diaminoalkylpyrrole, R- diaminoalkylimidazole, "b"- β-alanine, "|"- γ-linker, "+" - γ-NH3+ linker, "_" β-DP tail, "-"?NH(CH2)2OH tail, "-"-NH(CH3) tail23. The second polyamide molecule is colored red.

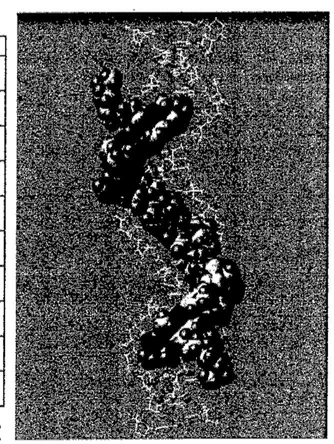


Fig. 5. Recognition of the target erbB2/Her2 DNA sequence 4 by the 8-ring tandem hairpin polyamide, predicted to have the best binding energy among ~300 polyamide designs tested (pairing diagram in the first row of Table

Diaminoalkyl moiety can interact with transcription factors in the DNA major groove and improve gene inhibition

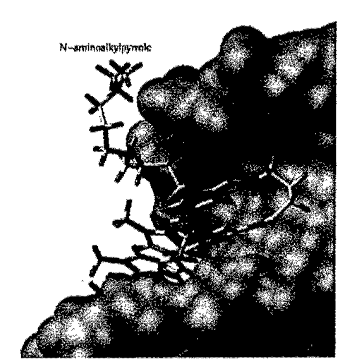


Fig. 6. N-diaminoalkylpyrrole containing polyamide in the DNA minor groove. This globally optimized conformation shows interaction of the diaminoalkyl tail with the DNA phosphates, which ensures inhibition of major-groove binding transcription factors by polyamides of this type.

References:
(1) (a) Wemmer, D. E., Annu Rev Biophys Biomol Struct 2000, 29, 439-61. (b) Dervan, P.B.; Burli, R. W., Curr Opin Chem Biol. 1999, 3, 688-93. (c) Gottesfeld, J.M., Turner, J.M., Dervan, P.B., Gene Expr 2000, 9, 77-9.
(2) (a) Abagyan, R.A.; Totrov, M.M.; Kuznetsov, D.A. J. Comp. Chem. 1994, 15, 488-506. (b) Katritch, V.; Abagyan R. A.; Olson, W. K. Abstracts of PMMB VI, Santa Fe, New Mexico 1999.
(c) Nemethy, G.; Gibson, K.D.; Palmer, K.A.; Yoon, C.N.; Paterlini, G.; Zagari, A.; Rumsey, S. & Scheraga, H.A. J. Phys. Chem. 1992, 96, 6472-6484.